# Cognitive Argumentation and the Selection Task

**Emmanuelle Dietz (emmanuelle.dietz@tu-dresden.de)**
International Center for Computational Logic, TU Dresden, Germany

**Antonis Kakas (antonis@ucy.ac.cy)**[1]
Department of Computer Science, University of Cyprus, Nicosia, Cyprus

## Abstract

This paper presents a study of the selection task based on Cognitive Argumentation (CA), a computational framework for dialectic argumentation-based reasoning. CA is built from a theoretical framework of argumentation in AI which is then grounded via cognitive principles from Cognitive Science. The aim is to understand the selection task variations by studying how argumentative reasoning is suitably flexible to uniformly capture the differences among the individuals' selections, the canonical groups and the shifts within the different contexts in which the experiment is carried out. Our approach is assessed with respect to the developed criteria within the meta-analysis of Ragni and Johnson-Laird in 2018.

**Keywords:** Selection Task, Conditional Reasoning, Cognitive Argumentation. Computational Modelling

## Introduction

Human Reasoning and Argumentation have been closely linked over the centuries, from as far back as Aristotle in Philosophy to recent works in Cognitive Psychology (Mercier & Sperber, 2011) and Artificial Intelligence (Dung, 1995). It is quite natural then to adopt the scientific hypothesis that "*Argumentation can form the basis for formulating human reasoning in a computational and cognitively adequate way*". In this paper, we examine this hypothesis using a framework, called **Cognitive Argumentation (CA)**, built by synthesizing together the general and abstract theory of computational argumentation from AI with cognitive principles born out of empirical and theoretical findings of Cognitive Science.

Dietz Saldanha and Kakas (2019) showed that human syllogistic reasoning (Khemlani & Johnson-Laird, 2012) can successfully be modeled within CA. Here, we will study how the same framework of CA, with the same formal structure and use of cognitive principles, is able to model the individual selections, the canonical groups and the variations within the selection task. The novelty of CA lies in the synthesis and bridge that it provides to exploit together advances from wide ranging studies in Philosophy, Cognitive Science and AI. Inferences in CA are cognitively valid through a mixture of descriptive cognitive principles, that are used to build the argumentation terrain and the weak normative requirement of acceptability of a coalition of arguments, as a self-defending team. Technically, CA is related to the mental model theory (Johnson-Laird, 1983), where now the structure corresponding to a model is that of an acceptable coalition of arguments understood as a *case* that adequately supports a conclusion. These cases are built, differently from mental models, via an underlying and uniform dialectic process of argumentative reasoning that follows a human-like (internal) debate of building arguments for and against a certain position.

In the original selection task (Wason, 1968), participants were presented with four cards, each card had a letter on one side and a number on the other side:

$$\boxed{D} \quad \boxed{F} \quad \boxed{3} \quad \boxed{7}$$

Which cards must be turned over in order to find out whether the following conditional holds? *If there is a D on one side of the card, then there is a 3 on the other side.* This task with abstract hypotheses, was reformulated by Griggs and Cox (1982) into a task with deontic principles: *If a person is drinking beer, then the person must be over 19 years of age.* They also carried out the task with everyday hypotheses (Wason & Shapiro, 1971), replicated similarly by Pollard (1981): *Every time I go to Manchester I travel by train.* Over the last decades many studies have tried to explain the participants' selections within the variations of the selection task. Ragni, Kola, and Johnson-Laird (2017; 2018) identified four canonical groups of participants as shown in Table 1.

Ragni and Johnson-Laird (2018) refuted 14 out of 16 theories by assessing them against the following criteria: (1) Can the theory predict the dependent selections of the individual? (2) Can the theory predict the canonical selections of the population? (3) Can the theory account for the changes of likelihood to select counter-examples when the task changed in framing or instruction? They propose the *new model theory*, which fulfills these criteria and is based the insight model (Johnson-Laird & Wason, 1970) and the mental model theory.

Guided by these criteria, we will develop an approach

[1]The authors are mentioned in alphabetical order.

Table 1: The four canonical selections and percentages for each group within the respective context. In the sequel, $P$, $Q$, $\overline{P}$, $\overline{Q}$ stand for $D$, 3, $F$, 7 (abstract), Manchester, train, Leeds, car (everyday), and beer, 22yrs, coke, 16yrs (deontic), respectively.

| Group | Selection | Abstract | Everyday | Deontic |
|---|---|---|---|---|
| I | $P$ | 36 | 23 | 13 |
| II | $P$, $Q$ | 39 | 37 | 19 |
| III | $P$, $Q$ and $\overline{Q}$ | 5 | 11 | 4 |
| IV | $P$, $\overline{Q}$ | 19 | 29 | 64 |

towards the selection task within Cognitive Argumentation (CA). We will introduce cognitive principles suitable to model conditional reasoning within CA. Through dialectic argumentation based reasoning we can model human decision to select a card or not. This allows us to characterize the selection task with respect to individual choices, the canonical groups and to investigate its relation to the new model theory.

## Cognitive Principles

Humans make assumptions while reasoning, many of which are not necessarily valid in formal logic. We will specify such typically extra-logical properties as *cognitive principles*. The formalization of such principles has previously been proposed to explain individual responses in syllogistic reasoning (Dietz Saldanha, Hölldobler, & Mörbitz, 2018; Dietz Saldanha & Schambach, 2020). Here, we will introduce those most relevant ones to reasoning in the selection task. They will guide us in the next section to formulate the framework of Cognitive Argumentation by suitably using them to shape the abstract framework of argumentation into a specific framework of argumentation suitable for cognitive human reasoning.

Reasoning in the selection task is primarily a case of conditional reasoning. We will also assume that hypothetical reasoning is involved, as already proposed by (Ragni et al., 2018), who emphasize the varying need for counterexample selection depending on the task setting. To further support this need for hypothetical reasoning we appeal to a cognitive principle of **Maxim of Relevance** (Grice, 1975). People consider different scenarios depending on whether they have been made *aware* of alternative options while reasoning (Sperber & Wilson, 1995; Byrne, 2005). This awareness may not be through some explicit mention of the alternative(s) but can also occur implicitly, resulting in the consideration of various context-dependent alternative hypotheses. This also ensures the existence of things and does not yield the paradox of confirmation as mentioned by Ragni et. al. (2018).

With respect to conditional reasoning we will assume that humans distinguish two types of relationship between a condition and a consequent: **Necessary** $\overset{n}{\rightsquigarrow}$ **and Sufficient Conditions** ($\overset{s}{\rightsquigarrow}$). Byrne (2005) distinguishes between such different types assumed to be perceived by humans. Consider the following two conditionals (Byrne, 1989): (1) *If she has an essay to finish, then she will study late in the library.* and (2) *If the library is open, then she will study late in the library.* In (1) The condition is sufficient whereas in (2) the condition is necessary. Similarly, in a conditional (3) *You must have a license, in order to drive.* the condition *you must have a license* is a necessary one.

Given a conditional, interpreting the condition as sufficient or necessary (or both) in relation to the consequent leads us to a directional association between the two propositions (of condition and consequent). This association can follow the sequence of the given conditional statement, as in the case of a sufficient condition, but for a necessary condition the association can be inverted, i.e. the necessary condition can appear as a consequence (principle of **Inverted Conditional**). For instance, we can reformulate (2) as follows: *If she will study late in the library, then the library is/ must be open.*. In such cases of a necessary condition we can form associations between the complements of the propositions, e.g. for the conditional (2) and (3): *If the library is not open, then she will not study late in the library.* and *If you do not have a licence, then you do/can not drive.*

Conditional reasoning is carried out under the evidence of some information we consider trustworthy and normally refer to it as factual information to indicate our confidence in the belief that it holds. According to Grice's (1975) conversational implicature, humans communicate according to a cooperation principle, namely the **Maxim of Quality** which states that the information that we get in conversation is assumed to hold and trusted. If a conflict arises between a fact and some other valid inference but which is based on some arbitrary or general assumption, e.g. a hypothesis, then this will be given up, in favor of the fact. We will encompass this as a principle of **Conflicts in Reasoning** through the adoption of the following relative strength among the various types of information associated to the cognitive principles above: Hypotheses are weaker than any other type of information (**Weakness of Hypothesis** principle) and the factual information is stronger than other conflicting information (**Strength of Facts** principle).

As a consequence of this relative strength principle we assume that people might recognize that a conditional statement together with some factual evidence can invalidate a hypothesis. If $P$ is a sufficient condition for $Q$ and hence we have a conditional association of *"If $P$ then $Q$"* ($\mathbf{suf}_1$), then under the factual evidence of $\overline{Q}$, people can recognize that the hypothesis of $P$ cannot hold, for if it did, this would support that $Q$ also holds which will be in direct conflict with the evidence $\overline{Q}$. Recognizing new associations of this type seems sometimes to be difficult because it requires an active search for counter-examples (O'Brien, D. S. Braine, & Yang, 1994; Rips, 1994). Yet, if this process is repeated several times, humans seem to form new associations, "short-cuts", that can be applied directly (principle of **learned**). In the example above they could make the association that from $\overline{Q}$, the contrary of $\overline{P}$ also follows, i.e. they form the additional **secondary or emergent** conditional association *"If $\overline{Q}$ then $\overline{P}$"* (**suf**$_2$) together with the **primary** one of **suf**$_1$. Analogously, primary and secondary associations with respect to necessary conditions can be specified (**nec**$_1$and **nec**$_2$). It is important to note that the classification of primary and secondary depends on the context of reasoning. The association that is straightforwardly formed because it is explicitly presented is primary, whereas the secondary association only emerges out of constructions of counter-examples and the defeat of their hypotheses.

## Cognitive Argumentation and Reasoning

The framework of Cognitive Argumentation (CA) is built by synthesizing cognitive principles within the framework of computational argumentation in AI, (Dung, 1995; Kakas & Moraitis, 2003; Modgil & Prakken, 2013). We briefly review the most relevant parts. The same argumentation framework from which CA is formed has been shown (Kakas, 2019) to capture Classical Propositional Logic in an equivalent way when the given premises are consistent and thus smoothly extend Classical Logic when the given theory is classically inconsistent. Also early on it was shown (see e.g. (Bondarenko, Dung, Kowalski, & Toni, 1997)) that most if not all AI non-monotonic logics could be reformulation in terms of argumentation.

The task is to understand the cognitive principles into a concrete computational form so that they can be reflected into computational argumentation. We review the basic components of the formal framework of CA. Details are found in (Dietz Saldanha and Kakas, 2019; 2020)

An **argumentation theory or model** within CA is obtained by translating cognitive principles into argument schemes together with a preference or strength relation between these schemes. It consists of a triple $\mathcal{A}_\mathcal{L} = \langle \mathcal{A}s, \mathcal{C}, \succ \rangle$ where $\mathcal{A}s$ is a set of argument schemes, $\mathcal{C}$ is a conflict relation in the language of the framework and $\succ$ is a binary strength relation on $\mathcal{A}s$. Argument schemes were introduced as stereotypical reasoning patterns that are typically non-deductive (Pollock, 1995; Walton, 1996). Formally, an **argument scheme**, $\mathsf{as} \in \mathcal{A}s$, is a tuple of the form $\mathsf{as} = (\mathsf{pre}, \mathsf{pos})$ where the **premises** $\mathsf{pre}$ and **position** $\mathsf{pos}$ are (sets of) statements in the language of discourse $\mathcal{L}$. Using an argument scheme $\mathsf{as} = (\mathsf{pre}, \mathsf{pos})$ we can construct an **individual argument** that **supports** the position $\mathsf{pos}$ based on the premises $\mathsf{pre}$. An **argument** $\Delta$**,** is then a set of individual arguments that are grouped together as a coalition to support a position (e.g. a conclusion) we are interested in. The **conflict relation** $\mathcal{C}$ in $\mathcal{A}_\mathcal{L} = \langle \mathcal{A}s, \mathcal{C}, \succ \rangle$ specifies when arguments conflict with each other and is used to give the notion of attack between two arguments. $\Delta'$ **attacks** or is a **counterargument** of $\Delta$, iff together these arguments have a conflict under contain $\mathcal{C}$, e.g. when $\Delta$ supports $Q$ and $\Delta'$ supports $\overline{Q}$, the negation or complement of $Q$. The **strength relation** $\succ$ in $\mathcal{A}_\mathcal{L} = \langle \mathcal{A}s, \mathcal{C}, \succ \rangle$ captures the relative strength among arguments. Given two argument schemes $\mathsf{as}$ and $\mathsf{as}'$, $\mathsf{as} \succ \mathsf{as}'$ means that arguments constructed from $\mathsf{as}$ are **stronger** than arguments constructed from $\mathsf{as}'$. This gives a notion of **defense** between conflicting arguments. Informally, an argument $\Delta$ defends against $\Delta'$ only when its individual arguments are at least as strong as those of $\Delta'$.

### Reasoning in Cognitive Argumentation

Reasoning in argumentation is a process of analysis of alternatives, e.g. of a conclusion and its negation, by a consideration of different arguments for and against the various competing alternatives. Three main requirements make a supporting argument of sufficiently good quality to be thus admitted in the frame of reasoning within CA.

This first one is **groundness** and relates to the observation that any instance of human reasoning is carried out under a current state of information that the environment gives or makes us aware of. This information is captured by a **cognitive state** $\mathcal{S} = (\mathcal{F}, \mathcal{A})$ where $\mathcal{F}$ is a *set of facts* provided by the environment and $\mathcal{A}$ is an *awareness set* of propositions which the environment has awaken as relevant in the current reasoning. An argument $\Delta$ is then required to be grounded on the current cognitive state, i.e. that all individual arguments in $\Delta$ need to eventually be based on information that refers to the cognitive state. The other two requirements are normally put together and specify a normative criterion of acceptability of arguments. Formally, an argument $\Delta$ is **acceptable or admissible** in $\mathcal{A}_\mathcal{L}(\mathcal{S})$ iff $\Delta$ is conflict-free (under $\mathcal{C}$) and $\Delta$ defends against all its counter-arguments attacking $\Delta$.

Reasoning under argumentation is then carried out by considering grounded and acceptable arguments that support a concluding statement $L$. This definition of acceptability is directly linked with the **dialectic argumentation process** of choosing a supporting argument, considering its counter-arguments and forming a defense against each one of these. Argumentation based reasoning is thus captured as a dialectic process along the following basic steps: (1) Construct a root argument supporting a conclusion of interest, (2) consider a counterargument against the root argument, (3) find a defense argument against the counterargument, (4) check that this defense argument is not in conflict with the root argument, (5) add this defense argument to the root argument, and repeat from (2), i.e. consider another counterargument to the now extended root argument.

In other words, argumentation based reasoning is a process of internal debate - contemplation - between supporting a position and trying to defeat it. Carrying out the process until there are no other counterarguments in step (2) that have not already being considered, clearly results in an extended root argument that is an acceptable argument supporting the conclusion of interest. We say that $L$ is **acceptable** or a **credulous conclusion** of $\mathcal{A}_{\mathcal{L}}(\mathcal{S})$, iff there exists a grounded and acceptable argument $\Delta$ in $\mathcal{A}_{\mathcal{L}}(\mathcal{S})$ that supports $L$. $L$ is a **skeptical conclusion** of $\mathcal{A}_{\mathcal{L}}(\mathcal{S})$ iff $L$ is a credulous conclusion of $\mathcal{A}_{\mathcal{L}}(\mathcal{S})$, and $\overline{L}$ is not a credulous conclusion of $\mathcal{A}_{\mathcal{L}}(\mathcal{S})$, i.e. there is no acceptable argument supporting $\overline{L}$. Credulous and skeptical conclusions represent **plausible** and **definite** conclusions, respectively. The existence of an acceptable argument supporting a conclusion makes this only a possible conclusion. Only when such possibilities of the contrary do not exist, we can derive in a definite conclusion.

The construction of acceptable arguments and the process of argumentation-based reasoning follows a simple **dialectic** form of (1) choosing a supporting argument (2) considering its counter-arguments and (3) forming a defense against each one of these. Details can be found in (Dung, Kowalski, & Toni, 2006).

## Argumentation in the Selection Task

We will now present how CA is applied to capture human reasoning in the selection task. We first need to ground the general framework of CA, according to the cognitive principles presented in the earlier section to give the specific framework of CA, denoted by $\mathcal{A}_{\mathrm{ST}} = \langle \mathcal{A}s, \mathcal{C}, \succ \rangle$, for the selection task. Each cognitive principle is translated into argument schemes in $\mathcal{A}s$ as follows.

The *maxim of quality* principle gives a **fact scheme**: $\mathsf{fact}(L) = (\emptyset, L) \in \mathcal{A}s$, applied for any statement $L \in \mathcal{F}$ of the current cognitive state $\mathcal{S} = (\mathcal{F}, \mathcal{A})$. Similarly, the *maxim of relevance* principle gives a **hypothesis scheme**: $\mathsf{hyp}(A) = (\emptyset, A) \in \mathcal{A}s$ and $\mathsf{hyp}(\overline{A}) = (\emptyset, \overline{A}) \in \mathcal{A}s$, applied for any proposition, $A$, in $\mathcal{S}$. The two different types of a condition $P$ in relation to a consequent $Q$, each give two **conditional** argument schemes: $P$ **is sufficient**: $\mathsf{suff}_1(P \rightsquigarrow Q) = (P, Q)$ and $\mathsf{suff}_2(\overline{Q} \rightsquigarrow \overline{P}) = (\overline{Q}, \overline{P})$, and $P$ **is necessary**: $\mathsf{necc}_1(Q \rightsquigarrow P) = (Q, P)$ and $\mathsf{necc}_2(\overline{P} \rightsquigarrow \overline{Q}) = (\overline{P}, \overline{Q})$

The conflict relation $\mathcal{C}$ is given by the negation between statements in the language and the strength relation $\succ$ as we have presented in the previous section, by: (1) hypotheses are weaker than any other opposing scheme and (2) facts are stronger than any other opposing scheme.

**Reasoning in the selection task** by any participant in the experiment will be done according to their interpretation of the condition in the conditional. This means that a person will reason using only a particular subset of the argument schemes in the framework $\mathcal{A}_{\mathrm{ST}}$. The argumentation reasoning process is then carried out by considering what acceptable arguments we can build for the information on the non-visible side of the card, given the factual information on the visible side. Specifically, we will consider whether it is possible to build acceptable argument supporting either the positive or negative statement of the other side of the card. The reasoner will then decide to turn a card if and only if **only** one of the two statements, positive or negative, on the non-visible side can be acceptably supported. In other words, **only** when the reasoner can **skeptically** conclude what is on the non-visible side of the card.

For each visible card, we will illustrate with figures the dialectic argumentation process for constructing acceptable/non-acceptable arguments supporting the two alternative conclusions on the other side of the card. In these figures, $\uparrow$ shows attacks and $\Uparrow$ shows *strong* defenses/attacks, i.e. that cannot be attacked back by the argument they are defending against/attacking.

**Card $P$.** Does $Q$ or $\overline{Q}$ follow? The cognitive state $\mathcal{S}$ is $(\{P\}, \{P, Q, \overline{P}, \overline{Q}\})$. If $P$ is understood as sufficient, then $\mathsf{suff}_1(P \rightsquigarrow Q) = (P, Q)$ is applicable and the following (strong) argument for $Q$ can be constructed:

$$\Delta_{P \overset{s}{\rightsquigarrow} Q} = \{\ \mathsf{fact}(P),\ \mathsf{suff}_1(P \rightsquigarrow Q)\ \}.$$

The only plausible counter-argument is the hypothesis that $Q$ does not hold: $\Delta^{\overline{Q}} = \{\mathsf{hyp}(\overline{Q})\}$. But this scheme is weak, i.e. $\mathsf{suff}_1(P \rightsquigarrow Q) \succ \mathsf{hyp}(\overline{Q})$. (preferences). Thus, $\Delta_{P \overset{s}{\rightsquigarrow} Q}$ is an acceptable argument, whereas $\Delta^{\overline{Q}}$ is not (Fig. 1). Therefore, $Q$ follows skeptically.

**Card $Q$.** Does $P$ or $\overline{P}$ follow? The cognitive state $\mathcal{S}$ is $(\{Q\}, \{P, Q, \overline{P}, \overline{Q}\})$. A (strong) argument for $P$ can be constructed, if $P$ is understood as a necessary condition:

$$\Delta_{Q \overset{n}{\rightsquigarrow} P} = \{\ \mathsf{fact}(Q),\ \mathsf{necc}_1(Q \rightsquigarrow P)\ \}.$$

The only plausible counter-argument is simply the hypothesis $\Delta^{\overline{P}} = \{\mathsf{hyp}(\overline{P})\}$. Again it holds that $\mathsf{necc}_1(Q \rightsquigarrow P) \succ \mathsf{hyp}(\overline{P})$. Accordingly, if $P$ is a necessary condition, $\Delta_{Q \overset{n}{\rightsquigarrow} P}$ is acceptable, whereas $\Delta^{\overline{P}}$ is not, and therefore $P$ follows skeptically (Fig. 2).

**Card $\overline{P}$.** None of the canonical groups contains this selection. It seems unlikely that $P$ is interpreted as necessary for $Q$ and even if, it seems difficult to admit $\mathsf{necc}_2(\overline{P} \rightsquigarrow \overline{Q})$. We are then left with the two, equally strong, hypothetical arguments $\Delta^{Q}$ and $\Delta^{\overline{Q}}$ which does not lead to a skeptical conclusion about $Q$.

**Card $\overline{Q}$.** Does $P$ or $\overline{P}$ follow? The cognitive state $\mathcal{S}$

Figure 1: Given $P$, construction for $Q$ (left) and for $\overline{Q}$ (right) assuming $P$ is a sufficient condition: $Q$ follows skeptically.

$$\begin{tikzcd}
\Delta_{P\overset{s}{\rightsquigarrow}Q} & \Delta_{P\overset{s}{\rightsquigarrow}Q} & \Delta_{P\overset{s}{\rightsquigarrow}Q} & \Delta^{\overline{Q}} & \Delta^{\overline{Q}} \\
& \Delta^{\overline{Q}} \arrow[u] & \Delta^{\overline{Q}} \arrow[u] & & \Delta_{P\overset{s}{\rightsquigarrow}Q} \arrow[u, Rightarrow] \\
& & \Delta_{P\overset{s}{\rightsquigarrow}Q} \arrow[u, Rightarrow] & &
\end{tikzcd}$$

Figure 2: Given $Q$, construction for $P$ (left) and $\overline{P}$ (right) assuming $P$ is a necessary condition: $P$ follows skeptically.

$$\begin{tikzcd}
\Delta_{Q\overset{n}{\rightsquigarrow}P} & \Delta_{Q\overset{n}{\rightsquigarrow}P} & \Delta_{Q\overset{n}{\rightsquigarrow}P} & \Delta^{\overline{P}} & \Delta^{\overline{P}} \\
& \Delta^{\overline{P}} \arrow[u] & \Delta^{\overline{P}} \arrow[u] & & \Delta_{Q\overset{n}{\rightsquigarrow}P} \arrow[u, Rightarrow] \\
& & \Delta_{Q\overset{n}{\rightsquigarrow}P} \arrow[u, Rightarrow] & &
\end{tikzcd}$$

Figure 3: Given $\overline{Q}$, construction for $\overline{P}$ (left) and for $P$ (right) assuming $P$ is a sufficient condition: $\overline{P}$ follows skeptically.

$$\begin{tikzcd}
\Delta_{\overline{Q}\overset{s}{\rightsquigarrow}\overline{P}} & \Delta_{\overline{Q}\overset{s}{\rightsquigarrow}\overline{P}} & \Delta_{\overline{Q}\overset{s}{\rightsquigarrow}\overline{P}} & \Delta^{P} & \Delta^{P} \\
& \Delta^{P} \arrow[u] & \Delta^{P} \arrow[u] & & \Delta_{\overline{Q}\overset{s}{\rightsquigarrow}\overline{P}} \arrow[u, Rightarrow] \\
& & \Delta_{\overline{Q}\overset{s}{\rightsquigarrow}\overline{P}} \arrow[u, Rightarrow] & &
\end{tikzcd}$$

is $(\{\overline{Q}\}, \{P, Q, \overline{Q}, \overline{P}\})$. If $P$ is understood as sufficient and $\mathsf{suff}_2(\overline{Q} \rightsquigarrow \overline{P})$ is recognized, the following (strong) argument for $\overline{P}$ can be constructed:

$$\Delta_{\overline{Q}\overset{s}{\rightsquigarrow}\overline{P}} = \{\ \mathsf{fact}(\overline{Q}),\ \mathsf{suff}_2(\overline{Q} \rightsquigarrow \overline{P})\ \}.$$

Its only counterargument $\Delta^P$ is defeated because $\mathsf{suff}_2(\overline{Q} \rightsquigarrow \overline{P}) \succ \mathsf{hyp}(P)$ (Fig. 3). Similarly, if the conditional is understood inverted and $Q$ as necessary for $P$ and $\mathsf{necc}_2(\overline{Q} \rightsquigarrow \overline{P}) = (\overline{Q}, \overline{P})$ is recognized we have a strong argument for $\overline{P}$. Otherwise, if $\mathsf{suff}_2(\overline{Q} \rightsquigarrow \overline{P})$ or $\mathsf{necc}_2(\overline{Q} \rightsquigarrow \overline{P}) = (\overline{Q}, \overline{P})$ are not recognized, we will only have the two weak but equally strong arguments $\Delta^P$ and $\Delta^{\overline{P}}$ that attack and defend each other: $P$ and $\overline{P}$ follow credulously, thus nothing follows skeptically. This lack of skeptical conclusion will occur similarly in any other of the above cards when the strong argument scheme involved is not recognized by the reasoner.

Summing up, each of the choice points in the dialectic argumentation process might non-deterministically be treated differently by different human reasoners, guided by the argument schemes they recognize.

## Characterization of the Canonical Groups

Table 2 summarizes the skeptical argumentative reasoning in the three contexts of the selection task. The third column shows the cognitive principles with the corresponding argument scheme that is applied and results, as illustrated in the previous section, in a skeptical conclusion. The last three columns show the dominant canonical selections, I, II and IV (selections from group III correspond to the ones from II and IV).

This allows us to link the canonical groups with the interpretation of the conditional as captured by the argument schemes and reasoning in CA. When reasoning with abstract and everyday hypotheses we have that, groups I and II do not seem to have the emergent associations established through counter-examples. Group I interprets $P$ as sufficient (for $Q$) only on the primary level, whereas II additionally understands $P$ as necessary, again only on the primary level. In contrast, for reasoning with deontic principles the **predominant** interpretation inverts the roles of the two statements in the given conditional, considering the consequent $Q$ to be necessary for its condition $P$. Interestingly, here the characterization for all groups is then analogous to the other contexts, but where the necessary and sufficient interpretations are swapped. Further, in all three contexts, groups I and IV involve only one interpretation of the condition, either only as sufficient, or only as necessary.

The gap between selecting $\overline{Q}$ from abstract (24%) to everyday (40%) hypotheses is explained by the learned principle as it makes secondary associations more emergent. The even higher percentage with deontic principles (68%) is explained by the observation that the condition within the inverted conditional is more easily interpreted as necessary both at primary and secondary level: The *16yrs* card easily produces a skeptical conclusion, a counter-example, for a high percentage of the population.

## New Model Theory Revisited

The new model theory (Ragni et al., 2018) adequately accounts for the individual selections, the canonical selections, and the variations wrt the recognition of counter-examples. Our approach and its resulting characterization of the canonical selection groups can be seen to share characteristics with the new model theory. Firstly, both approaches are based on inference with the conditional rather than relying on some meaning of the conditional. Secondly, our levels of reasoning with primary or secondary argument schemes is analogous to the levels of insight and the way these are used in the new model theory. Partial and complete insight in the new model theory are related to reasoning with counter-examples. Our reasoning with the hypothetical argument scheme is analogous to reasoning with counter-examples. This results in a corresponding understanding of many selections: Consider group IV when reasoning with abstract and every-

| | Card | % | Cognitive Principles | I | II | IV |
|---|---|---|---|---|---|---|
| Abstract | D | ≥99 | $\mathsf{suff}_1(P \rightsquigarrow Q)$ | ✓ | ✓ | ✓ |
| | 3 | 44 | $\mathsf{necc}_1(Q \rightsquigarrow P)$ | | ✓ | |
| | 7 | 24 | $\mathsf{suff}_2(\overline{Q} \rightsquigarrow \overline{P})$ | | | ✓ |
| Everyday | Manchester | 100 | $\mathsf{suff}_1(P \rightsquigarrow Q)$ | ✓ | ✓ | ✓ |
| | Train | 48 | $\mathsf{necc}_1(Q \rightsquigarrow P)$ | | ✓ | |
| | Car | 40 | learned $\mathsf{suff}_2(\overline{Q} \rightsquigarrow \overline{P})$ | | | ✓ |
| Deontic | Beer | 100 | inverted $\mathsf{necc}_1(P \rightsquigarrow Q)$ | ✓ | ✓ | ✓ |
| | 22 yrs | 23 | inverted $\mathsf{suff}_1(Q \rightsquigarrow P)$ | | ✓ | |
| | 16 yrs | 68 | inverted learned $\mathsf{necc}_2(\overline{P} \rightsquigarrow \overline{Q})$ | | | ✓ |

Table 2: Schemes and selections for Group I, II and IV.

day hypotheses the selection of $\overline{Q}$ is similarly explained by the application of partial insight in the new model theory and the use of the secondary sufficient scheme in CA.

Only two theories seem to adequately model the selection (Ragni et al., 2018): The model theory which relies on the *meaning* of a hypothesis and the inference-guessing theory which relies on *inferences* from the hypothesis. CA relies on both, the interpretation of the hypothesis within the content of the task guided by the underlying principles, and skeptical inference through argumentation. Nevertheless, the different perspective of argumentation-based reasoning in our approach allows the possibility of new explanations of the data. For example, the significant increase in the percentage when reasoning with deontic principles in Group IV (68%), can now be understood to result by interpreting the consequent of the given conditional inversely and as necessary for the condition in the given conditional. At the same time this inverted and necessary interpretation explains the drop of the percentages of Group I and II compared to the other two contexts. We are also able to establish a new analogy between the three contexts: the interpretation of necessary and sufficient conditions is swapped when reasoning with deontic principles.

There is only one free parameter in CA and that is the choice of argument schemes admitted at the time of reasoning. This then reflects also the intensity or thoroughness of the reasoning by the human reasoner. This superiority is also supported by the universal encapsulation/reformulation of all AI non-monotonic logics in terms of argumentation (Bondarenko et al., 1997).

The criticism on Ragni et. al. work by Kellen and Klauer (2020) does does not affect our approach. We address three of their arguments: (i) Ragni et. al. excluded a priori several selections as potentially canonical, such as (only) $q$. This seems reasonable, even though we understand Kellen et. al., no selection should be excluded a priori. CA could account for this special case with a simple explanation: Participant do not understand the condition as sufficient. (ii) Dependencies at the individual-person level cannot be inferred from aggregated data. We only considered non-aggregated data. (iii) The amount of parameters is not a good proxy for a good model. This argument does not hold for CA as its strength is its simple and general. Independent of Kellen et al.'s criticism, we still believe that Ragni et. al.'s benchmarking efforts will have a significant value for the community.

### COGNICA and Future Directions

The application of CA to conditional reasoning has been implemented in a system called COGNICA. Currently, we are extending COGNICA to admit a wider variety of conditionals as studied in the work of Laird and Byrne (2002) and to reason with knowledge that can include together different types of conditionals. We are designing experiments that are aimed to test directly the underlying assumptions of the framework of CA. In particular, we are planning to empirically test how explanations, that are naturally connected with argumentative reasoning, can affect the human participants' reasoning and to examine if the COGNICA system's explanations can lead them to reconsider their inferences and/or the confidence they have in their drawn inferences.

## Conclusions

Based on the hypothesis that argumentation is a good basis for modeling human reasoning we have shown that the interdisciplinary framework of CA can capture the data in the selection task, in a simple and unified way, across all the canonical groups and changes in these over the different contexts of the task. Technically, in CA the variation in the data is captured through one parameter, that of the subset of argument schemes admitted in the reasoning, under the same and singular criterion of acceptability of grounded arguments. Our approach accords with the new model theory of (Ragni et al., 2018): argumentation-based reasoning is performed as a constructive dialectic process where the subset of argument schemes used by a reasoner reflects the potential computational intensity of the reasoning, analogously to the level of insight of the new model theory.

The main advantages of CA is its simplicity, its generalizability and its use of a universal criterion of acceptability. We believe that this work between AI, Explainability and Cognitive Science will strengthen the already existing link between argumentation and reasoning.

## References

Bondarenko, A., Dung, P. M., Kowalski, R. A., & Toni, F. (1997). An abstract, argumentation-theoretic approach to default reasoning. *Artif. Intell.*, *93*, 63–101.

Byrne, R. M. J. (1989). Suppressing valid inferences with conditionals. *J. of Memory and Language*, *31*.

Byrne, R. M. J. (2005). *The rational imagination: How people create alternatives to reality*. MIT press.

Dietz Saldanha, E.-A., Hölldobler, S., & Mörbitz, R. (2018). The syllogistic reasoning task: Reasoning principles and heuristic strategies in modeling human clusters. In D. Seipel, M. Hanus, & S. Abreu (Eds.), *Declarative programming and knowledge management* (Vol. 10997).

Dietz Saldanha, E.-A., & Kakas, A. C. (2020). Cognitive argumentation and the suppression task. *CoRR*, *abs/2002.10149*.

Dietz Saldanha, E.-A., & Schambach, R. (2020). A computational approach for predicting individuals' response patterns in human syllogistic reasoning. In S. Denison, M. Mack, Y. Xu, & B. C. Armstrong (Eds.), *Proc. of CogSci 2020.*

Dietz Saldanha, E.-A., & Kakas, A. (2019). Cognitive argumentation for human syllogistic reasoning. *KI - Künstliche Intelligenz*, *33*(3), 229–242.

Dung, P. M. (1995). On the Acceptability of Arguments and its Fundamental Role in Nonmonotonic Reasoning, Logic Programming and n-person Games. *AI*, *77*.

Dung, P. M., Kowalski, R. A., & Toni, F. (2006). Dialectic proof procedures for assumption-based, admissible argumentation. *AI*, *170*(2).

Grice, H. P. (1975). Logic and conversation. In P. Cole & J. L. Morgan (Eds.), *Syntax and semantics* (Vol. 3). New York: Academic Press.

Griggs, R. A., & Cox, J. R. (1982). The elusive thematic-materials effect in wason's selection task. *British J. of Psychology*, *73*(3), 407–420.

Johnson-Laird, P. (1983). *Mental models: towards a cognitive science of language, inference, and consciousness*. Harvard University Press.

Johnson-Laird, P., & Wason, P. (1970). A theoretical analysis of insight into a reasoning task. *Cognitive Psychology*, *1*(2), 134 - 148.

Johnson-laird, P. N., & Byrne, R. M. J. (2002). Conditionals: A theory of meaning, pragmatics, and inference. *Psychological Review*, 646–678.

Kakas, A. C. (2019). Informalizing formal logic. *Informal Logic*, *39*(2), 169–204.

Kakas, A. C., & Moraitis, P. (2003). Argumentation based decision making for autonomous agents. In *Proc. of 2nd int. joint conf. on autonomous agents & multiagent systems, AAMAS* (pp. 883–890). ACM.

Kellen, D., & Klauer, K. (2020). Modeling the wason selection task: A response to Ragni and Johnson-Laird (2020) doi: https://doi.org/10.31234/osf.io/uvmqc. *PsyArXiv*.

Khemlani, S., & Johnson-Laird, P. N. (2012). Theories of the syllogism: A meta-analysis. *Psychological Bulletin*, *138*(3), 427-457.

Mercier, H., & Sperber, D. (2011). Why do humans reason? arguments for an argumentative theory. *Behavioral and Brain Sciences*, *34*(2), 57-74.

Modgil, S., & Prakken, H. (2013). A general account of argumentation with preferences. *AI*, *195*, 361–397.

O'Brien, D., D. S. Braine, M., & Yang, Y. (1994, 11). Propositional reasoning by mental models? simple to refute in principle and in practice. *Psych. review*, *101*.

Pollard, P. (1981). The effect of thematic content on the 'wason selection task'. *Current Psychology*, *1*(1).

Pollock, J. L. (1995). *Cognitive carpentry: A blueprint for how to build a person*. MIT Press.

Ragni, M., Kola, I., & Johnson-Laird, P. (2017). The wason selection task: A meta-analysis. In G. Gunzelmann, A. Howes, T. Tenbrink, & E. Davelaar (Eds.), *Proc. of CogSci 2017.* Cognitive Science Society.

Ragni, M., Kola, I., & Johnson-Laird, P. N. (2018). On selecting evidence to test hypotheses: A theory of selection tasks. *Psychological Bulletin*, *144*(8).

Rips, L. J. (1994). *The psychology of proof: Deductive reasoning in human thinking*. The MIT Press.

Sperber, D., & Wilson, D. (1995). *Relevance: Communication and cognition*. Oxford: Blackwell Publishers.

Walton, D. N. (1996). *Argumentation schemes for presumptive reasoning*. LEA.

Wason, P. C. (1968). Reasoning about a rule. *Quarterly J. of Experimental Psychology*, *20*(3), 273–281.

Wason, P. C., & Shapiro, D. (1971). Natural and contrived experience in a reasoning problem. *Quarterly Journal of Experimental Psychology*, *23*(1), 63-71.